Data Management for the
Salton Sea Restoration Monitoring and Assessment Plan
California Department of Fish & Game
March 25, 2008

Introduction

The overarching goal of the Salton Sea Restoration Monitoring and Assessment Plan (MAP) is to implement a data collection, analysis, management, and reporting system to inform and guide management actions for the restoration of the Salton Sea ecosystem. Various workgroups were formed to address the diverse aspects of the MAP. The following focused technical workgroups were created: Air Quality and Climatological, Biological, Hydrological, Geographic/Geologic, Socioeconomic, and Data Management. The Data Management workgroup will be lead by the California Department of Fish and Game (DFG). The purpose of this document is to discuss the Data Management Plan for the MAP.

Objectives

It was determined during the initial MAP meetings that Data Management would need to meet the following objectives:

1. Define contract and budget expectations for Data Management.
2. Establish and enforce standards for data documentation.
3. Establish data transfer and storage protocol.
4. Make data easily accessible and in a useable format.
5. Perform a retrospective analysis on existing data sets.

Objective 1: Define contract and budget expectations for Data Management

- Data deliverables and metadata standards must be implicitly stated in project planning and contracts. All contracts need to contain the following verbiage:

  "The spatial data collected and created for this contract is a required deliverable of this contract and will become the property of the Department of Fish and Game, and not of the contracted party. A condition of final payment on this contract shall include the delivery of all related spatial data (in an ESRI useable format where applicable), documented with metadata in accordance with minimum BIOS metadata standards (http://bios.dfg.ca.gov/metadata.asp) and FGDC metadata standards (http://www.fgdc.gov/metadata/documents/workbook_0501_bmk.pdf)."

- It is important to stress that all deliverables be submitted during the funding period; otherwise, "knowledge, staff and motivation may be lost and the work may not be carried out to adequate standards" (Guidance on Data Management 18).

- Database design must be reviewed by the Data Management team to ascertain compatibility in BIOS and ensure metadata has been started. Metadata is best started during data design where questions such as purpose, objective, methods, field definitions, etc. can be answered. See Appendix 1 for an outline of questions that can be answered during the data design process. Answering these questions early in a project will result in complete and accurate metadata by the projects completion.
- Project budgets should allocate funds for management, dissemination, and archiving of data. "Data Management costs can exceed those required for initial data collection. Independent studies have published estimates of a minimum percentage of project funds that should be allocated to information management. These estimates range from 10% (NRC 1995) to 18% (GCRP 1992)" (Landis and Palmer 14).

Objective 2: Establish and enforce standards for data documentation

- All deliverable products must contain metadata. The Federal Geographic Data Committee (FGDC) defines metadata as "*'data about data' [sic] They describe the content, quality, condition, and other characteristics of data. Metadata help a person to locate and understand data"* (Content Standard 1). In order to guarantee metadata as a deliverable product contracts should contain the following text:

  > "The spatial data collected and created for this contract is a required deliverable of this contract and will become the property of the Department of Fish and Game, and not of the contracted party. A condition of final payment on this contract shall include the delivery of all related spatial data (in an ESRI useable format where applicable), documented with metadata in accordance with minimum BIOS metadata standards (*http://bios.dfg.ca.gov/metadata.asp*) and FGDC metadata standards (*http://www.fgdc.gov/metadata/documents/workbook_0501_bmk.pdf*)."

- Metadata must meet DFG's metadata standards as defined in Appendix 2.
- Metadata must meet the FGDC guidelines as stated in the *Content Standard for Digital Geospatial Metadata Workbook* found at: *http://www.fgdc.gov/metadata/documents/workbook_0501_bmk.pdf* .
- Data collection methods and standards must be documented and followed consistently throughout a project. "The *best time to collect metadata* is while the data are being developed, when the information needed for metadata is known. Waiting until after the data are developed risks less accurate information being recorded and increased costs caused by searching for information" (Content Standard 8). "Finalizing metadata is best done as soon after data collection as possible" (Landis and Palmer 7).

Objective 3: Establish data transfer and storage protocol

Focused Technical Group's Role and Responsibilities:

Before Objective 3 can be formalized the following questions must be addressed by the various Focused Technical Groups:

1. What is the anticipated storage size of the data?
2. What is the anticipated transfer frequency?
3. How have individuals accessed their data in the past? How have they summarized their data for internal and external needs in the past? How do they envision their data being displayed? The answer to these questions will determine how we will manage and display data with a fixed geographic location with data being collected at fixed time intervals such as air quality data.
4. What data is anticipated to be stored at DFG and what data will be stored with other agencies and where? For example, will the California Air Resources Board (CARB) maintain air quality data? Will the State Water Resources Control Board's (WRB) Surface Water Ambient Monitoring Program (SWAMP) maintain water quality data? Do these other agencies anticipate designing and implementing an ArcGIS server application?

- Data will be delivered in a digital format. Any non-digital data that is collected, such as paper forms, will need to be converted to a digital format.
- Data must be delivered in an acceptable and usable format. GIS data must be delivered in an ESRI format. Databases must be in either Access or Excel. Word documents must be in a Microsoft format or in a searchable PDF format. Word documents must be formatted to be accessible to people with disabilities.
- Transfer of data to DFG will be based on size. Datasets smaller than 10 MB can be transferred through email to kwhite@dfg.ca.gov. Datasets larger than 10 MB can be placed on DVD and mailed to Kristina White at 1807 13th Street Sacramento, CA 95811 or available portable USB drives can be borrowed from DFG.
- Focused Technical Groups (FTG) must consider the needs of both the public and the technical user when compiling, assembling, and disseminating data. For example, the public prefers to see data summarized into reports, maps, visual aids, figures, etc.; whereas, technical users prefer to see the raw datasets, technical articles, manuals, protocols, etc.

DFG's Role and Responsibilities:

- Outside requests for data will be based on information contained in the metadata's use and access sections. If a dataset is distributable then transfer of the data will be based on size and the delivery method will be the same as that mentioned in the transfer of data section above. If a dataset cannot be redistributed then the individual will be directed to the point of contact listed in the metadata.

- Data is backed-up based on the rate of changes made. This can mean that data is backed up every day or once a month. These backups are stored on tape which is taken offsite once a month.
- Versioning will be documented within the data quality section of the metadata and dataset names will contain date stamps.
- Access control will be enforced by securing the dataset location and access will only be provided to authorized personnel.

Objective 4: Make data easily accessible and in a useable format

- Access to geographic data that is collected for the Salton Sea MAP would be through DFG's Biogeographic Information and Observation System (BIOS) internet map viewer (http://bios.dfg.ca.gov). A limitation of the current BIOS map viewer is that it is running on ArcIMS technology, which does not allow for easy integration of data from outside sources. However, DFG is working on updating BIOS to ArcGIS Server technology which would allow for this integration. The goal is to make BIOS able to bring in specialized data being housed in other Departments (like Air Quality and Water Quality data), and display it along with all of the other Salton Sea data in a single map viewer.
- Metadata for spatial and non-spatial data will be distributed to metadata clearinghouses to provide access to a wider audience. Examples of metadata clearinghouses include, but are not limited to, the National Biological Information Infrastructure (NBII), California Environmental Resources Evaluation System (CERES), and the National Spatial Data Infrastructure (NSDI).
- Data that does not have a geographic component, does not need to be viewed from the BIOS map viewer, and does not need to be searched by a wider audience will be documented and stored in DFG's Document Library. The Document Library provides a single storage area for a variety of documents, and is available to the public through the internet. The Document Library provides an easy way to search through all of the stored documents to find a particular topic. Depending on the need, each individual document can be made public or can be secured so that only specific users can view it.
- To ensure compatibility among diverse users, standard data definitions, terminology, and software formats should be used. The Data Dictionary maintained by DFG contains searchable documentation and/or downloads for database diagrams, glossaries, reference (or lookup) tables, business rules, protocols and schema definitions. For example, in the Reference Table section you can download "look up" table data to use in your own work or you can search for the definition of a code. New reference tables can be created to suit the different monitoring projects. Reference tables can also be modified and updated based on user input.
- Where applicable, derived data products such as reports, summaries, graphs, etc. should be provided to DFG. These products are often more useful to the end-user than the raw unprocessed data (Landis and Palmer 2).
- Data should be delivered in a useable format and include information on data-collection methods and quality of data (Palmer 215). Examples of useable

formats include ESRI formatted GIS data, Microsoft office formats such as Access, Word, Excel, etc. (see Appendix 3)

Objective 5: Perform a retrospective analysis on existing data sets

- Any data identified by the heads of each FTG needs to be communicated back to the Data Management team for cataloging and potential acquisition based on prioritization.
- Existing data must be prioritized "with regards to the effort and costs required for retrieving, reformatting, cataloguing, and re-archiving it versus the 'value' of the data or information" (Landis and Palmer 13). Prioritization should be based on the quality of the data and its documentation. If data lacks documentation or if documentation is scarce then the value of the dataset is diminished. "Documenting existing or 'legacy' data can be daunting. Details are long forgotten and costs can be high" (Content Standard 8).
- Existing data and data sources can be documented and searched using DFG's Document Library, or if the value of the dataset is high it can be placed in BIOS.
- Specific data being displayed on the Redlands Institute Salton Sea Map Viewer (http://www.institute.redlands.edu/salton/data/maps.aspx) can be incorporated into BIOS by request from the various FTG's.
- Known existing data sources include, but are not limited to, Torres Martinez, CARB, Redlands Institute, CH2M, Scripps Institution of Oceanography, Coachella Valley Association of Governments (CVAG), Imperial Irrigation District (IID), Riverside County, Imperial County, California Department of Water Resources (DWR), DFG, and WRB.

BIBLIOGRAPHY

Content Standard for Digital Geospatial Metadata Workbook – Version 2.0. 1 May 2000. Online. 30 January 2008. Available WWW: http://www.fgdc.gov/metadata/documents/workbook_0501_bmk.pdf

GCRP, 1992. USGC Data and Information Program Plan. U.S. Global Change Research Program. Committee on Environmental and Natural Resources, National Science and Technology Council, Washington, D.C.

Guidance on Data Management. August 2006. Online. 7 January 2008. Available WWW: http://www.data-archive.ac.uk/relu/reluaug2006.pdf

Landis, Eric B. and Craig J. Palmer. Lake Mead Recreation Area Natural Resources Division: Data and Information Management Procedures. 10 January 2002. Online. 7 January 2008. Available WWW: http://hrcweb.lv-hrc.nevada.edu/mojn/data/DMP_LAME.pdf

NRC, 1995. "Finding the forest in the trees: The challenge of combining diverse environmental data," National Academy Press, Washington, D.C. 129 pp.

Resources for Implementing the WWF Project and Programme Standards. May 2006. Online. 13 February 2008. Available WWW: http://assets.panda.org/downloads/4_1_manage_incoming_data_05_17_06.pdf

Palmer, Craig J. "Approaches to Quality Assurance and Information Management for Regional Ecological Monitoring Programs." Monitoring Ecosystems: Interdisciplinary Approaches for Evaluating Ecoregional Initiatives. Washington, D.C.: Island Press, 2003: 211-255.

Appendices

1. Resource Management Project Plan Outline for Metadata Collection
2. Minimal DFG Required Metadata
3. Standards That Enable You to Contribute Your Data to Larger Data Sets

Appendix 1: Resource Management Project Plan Outline for Metadata Collection (taken from Landis and Palmer 18-19)

Title: What is the title of the project?

Team: Who is the project lead and who are the project participants?

Background:

Purpose: What is the purpose of the project?
Objectives: What are the project objectives?
Audience: Who is the principal audience for the project?
Related Work: List any related data sets or reports that could be documented for cross-reference.

Approach:

Collection: Will the data be developed primarily through:
a) Field visits?
b) Remote instrumentation (i.e. temperature recorders, etc)?
c) Existing data sources (please list)?
Design: Where will the data be collected?
Timeline: What is the anticipated time period in which the data will be collected?
Methods: Briefly summarize your field and laboratory methods (cut & paste from other documents! If you used existing protocols or methods, list the references).
Species: What species or communities will be examined?
a) Will you be collecting any specimens and removing them from the site?
b) If so, what, how many and what will be their disposition?
Taxonomy: Will you use a taxonomic authority or field guide for identification? If so, what is the reference?
Exclusion: Will you exclude anything from your data collection? (i.e. stems less that a certain diameter or streams without surface flow)

Results:

Use: How will the collected data be used?
Publishing: Will the data set (or resulting analysis) be published or part of a larger publication? If so, what will the reference be?
Summary: How do you plan to summarize your data?
Models: Will you use a model or other analytical tool to develop your dataset? If so, what is the reference?
Restrictions: Are there legal restrictions on who may use the data?

Data Management:

Owner: Who will be the originator(s)/ owner of the data collected during the project? (Include address and telephone number)

a) If someone else will collect data, please list the name(s), address, and telephone number.
b) Are there other organizations or individuals who should get credit for support, funding, or data collection and analysis?

Form: What will be the form of your data set? - (spreadsheet, ascii file, gis layer, database, other) Why this form?

Filename: What will be the filename for your data set?

Fields: For each file or table, list the fields in the data set and for each field list:
a) The definition of the field.
b) If the data will be coded, list the codes and the definitions.
c) If the codes come from a published code set, list the reference.
d) If the data are measured, list the units and the minimum and maximum allowable values.
e) Otherwise, the domain is unrepresentable. Include a brief description of what is in the field.

Collection sheets: Please attach a copy of all draft data collection sheets.

Updates: Will the data set be updated? If so, how frequently?

Archive: Where will your data set be archived (short-term and permanent)?

Keywords: List some keywords to help search for this data set?

Advice: Do you have any advice for potential users of the data set?

Distribution: What are your distribution instructions, if any?

GIS Data: Will this be a GIS data set? If no, skip to next question.
a) What are the projection parameters?
b) List the processing steps you will use to create your data set, including the approximate date of processing.

Quality Assurance:

MQO's: What are your measurement quality objectives for each parameter of measurement? (If these are included with the methods, simply refer to the methods section)

Training: What is your plan for training and certifying data collection staff?

Audits: When do you plan to audit your field crews? If available, attach field audit form.

Data Checking: What measures will you take to make certain that your data set is as nearly correct as possible? (i.e.- verification/validation)

Quality Assessment: How do plan to assess the quality of your measurements?

Appendix 2: Minimal DFG Required Metadata

Please provide these items for each dataset.

Abstract: Briefly describe what the data set is about (who, what, where, when). Include any limitations of the dataset, assumptions made, and if there is anything special that the user of these data should be aware of.

Purpose: Briefly describe why the data set was created.

Date: The date or range of dates when the data were gathered, or the date the photos, maps or other items at the core of the data set, were created.

Point of Contact: Contact information for an individual or organization that is knowledgeable about the data set. Include:

Person's Name: Complete first and last name
Organization's Name: Program, administrative unit, and agency, company, or group name
Telephone Number: Including Area Code
E-Mail address:

Field Definitions: List and define each field used in your shapefile, database, or spreadsheet.

Abbreviation Definitions: For any field that contains numeric or alphabetic codes (e.g., SAC = Sacramento County), list each code/abbreviation and provide an unabbreviated definition.

Access Constraints: Is there a need to limit who has access to see or read this dataset? If so, specify. If not, put "**None**".

Use Constraints: Is there a need to limit the use of this dataset to certain people or to specific tasks? If so, specify, If not, put "**None**". Also include how the data should be cited, if you want something specific.

Data Distribution: Can your data be distributed? If yes, to who?

Progress: **Complete** or **Incomplete**.

Update Frequency: Possible values are: **Continually, Daily, Weekly, Monthly, Annually, Unknown, As Needed, Irregular, None Planned**, or …..

If you are providing a GIS file (shapefile, etc), these next 2 items are often defined as part of that file, but you'll need to make sure.

Projection: What is the Projected Coordinate System name?
California Teale Albers (*preferred*)
Latitude/Longitude
State Plane
UTM

Datum (or Geographic Coordinate System): Which Datum is the projection in?

NAD83	**(GCS_North_American_1983) (*preferred*)**
NAD27	**(GCS_North_American_1927)**
WGS84	**(WGS_1984)**

Keywords (optional): Words or short phrases summarizing an aspect of the data set, used to allow people to find your dataset with quick keyword searches.

Theme	Subjects covered by the data set.
Place	Geographic locations characterized by the data set.

Appendix 3: Standards That Enable You to Contribute Your Data to Larger Data Sets (taken from Resources for Implementing the WWF Project and Programme Standards 6)

One of the most important places you can contribute your data are to the growing number of databases that are developing at national, regional, and global levels around the world. If you contribute your data, then other practitioners can make use of your findings and learn from your experiences. In order to make your data accessible to outside parties, they need to conform to international data standards that provide the basis for open sharing of data. These standards need to occur on several levels:

Standard Software Formats – Your data need to be in an electronic format that other users can either directly read or at least import. For example, because of their dominant market position, most people can read files generated by Microsoft programs that use the *.doc, *.xls and *.mdb formats for documents, spreadsheets and databases respectively. Non-proprietary formats include HyperText Markup Language (HTML) used for web-pages or, in more recent years, various flavours of eXtensible Mark-up Language (XML). For example, in the world of spatial data, ESRI's Geography Markup Language (GML) is becoming the standard XML encoding for geospatial information.

Standard Data – Your data also need to fit the structure of the databases that you will be contributing to. As a simple example, if you are reporting numbers of birds of different species in a census and your data are in the form of nesting pairs whereas the database wants to know individual adult birds, your data will not be compatible unless they are converted. Most databases will outline the format that data need to be in.

Standard Terms – A particularly important aspect of the need for standard data is the need to have standard terminology. If you have recorded bird names in the local language, chances are they won't be useful at a global level. To this end, it's important to use scientific (latin) names. Similarly, if you call a threat "cattle grazing" and another project terms it "livestock" then there will be no way to compare results. Specific resources that you may wish to consult for terminology include:

• Geographic Place Names – GeoNet Names Server (GNS): http://gnswww.nga.mil/geonames/GNS/index.jsp

• Biological Species and Other Taxonomic Information – Integrated Taxonomic Information System (ITIS): http://www.itis.gov/

• Habitats, Threats, and Conservation Actions – IUCN/CMP Classifications & Authority Files: http://www.redlist.org/info/authority_files

• General Terms: California Environmental Resources Evaluation System (CERES): http://gis.ca.gov/catalog/thesaurus.epl?mode=browse or General Multilingual Environmental Thesaurus (GEMET): http://www.eionet.europa.eu/gemet

Standard Metadata – In addition to contributing your data themselves, you should also contribute meta information about your dataset. Typical metadata include identification

of the data being described, the source of the data and a contact person/organization, the quality of the data, entity, or attribute information (if in a database or spreadsheet), its publication date, distribution information (including rights/liabilities), and the name of the individual completing the metadata. Metadata are typically recorded in a separate file (often using HTML or XML) that accompanies the main data file. The US Federal Geographic Data Committee has created a metadata standard that is fairly widely accepted throughout the world and is suggested as a good metadata reference to employ. For more information on this standard, please go to: http://www.fgdc.gov/metadata/geospatial-metadata-standards.